Long-term variations in the magnetic fields of the Sun and the heliosphere: their origin, effects, and implications

Article

Published Version

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 106, NO. A8, PAGES 16,021–16,038, AUGUST 1, 2001

Long-term variations in the magnetic fields of the Sun and the heliosphere: Their origin, effects, and implications

M. Lockwood [1]

Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, England, UK

Abstract. Recent studies of the variation of geomagnetic activity over the past 140 years have quantified the "coronal source" magnetic flux F_s that leaves the solar atmosphere and enters the heliosphere and have shown that it has risen, on average, by an estimated 34% since 1963 and by 140% since 1900. This variation of open solar flux has been reproduced by *Solanki et al.* [2000] using a model which demonstrates how the open flux accumulates and decays, depending on the rate of flux emergence in active regions and on the length of the solar cycle. We here use a new technique to evaluate solar cycle length and find that it does vary in association with the rate of change of F_s in the way predicted. The long-term variation of the rate of flux emergence is found to be very similar in form to that in F_s, which may offer a potential explanation of why F_s appears to be a useful proxy for extrapolating solar total irradiance back in time. We also find that most of the variation of cosmic ray fluxes incident on Earth is explained by the strength of the heliospheric field (quantified by F_s) and use observations of the abundance of the isotope ^{10}Be (produced by cosmic rays and deposited in ice sheets) to study the decrease in F_s during the Maunder minimum. The interior motions at the base of the convection zone, where the solar dynamo is probably located, have recently been revealed using the helioseismology technique and found to exhibit a 1.3-year oscillation. This periodicity is here reported in observations of the interplanetary magnetic field and geomagnetic activity but is only present after 1940. When present, it shows a strong 22-year variation, peaking near the maximum of even-numbered sunspot cycles and showing minima at the peaks of odd-numbered cycles. We discuss the implications of these long-term solar and heliospheric variations for Earth's environment.

1. Introduction

1.1. Solar Cycle and Secular Variations

There has been a growing interest in long-term variations of the Sun. What was once termed the "solar constant" has been shown to vary with the sunspot cycle and thus is now generally referred to as the "total solar irradiance", S. The amplitude of this variation is of order 0.1% [*Willson*, 1997; *Fröhlich and Lean*, 1998]. Irradiance variations are mainly caused by the combined effect of darkening by sunspots with the brightening of associated faculae and of the network. Of the two, the brightening is the dominant effect [*Chapman et al.*, 1997], but both are associated with the magnetic field at the solar surface [*Fligge et al.*, 1998]. Variations of the irradiance on this 11-year timescale have not been considered to be greatly significant for Earth's climate because their effects are smoothed by the heat capacity of the oceans [*Wigley and Raper*, 1990]. However, continuous and homogeneous sunspot observations have been made since 1749, and sunspot numbers show longer-term, secular variations (on timescales of order 100 years and greater) in addition to the 11-year solar activity cycle [*Gleissberg*, 1944; *Pulkkinen et al.*, 2000]. Any associated changes in the total solar irradiance on these longer timescales would be highly significant for global climate change because they would not be smoothed. Thus an understanding of the ratio of the amplitudes of the variations on 11-year and longer timescales is of great importance. Using various proxy data, attempts have been made to extrapolate the recently observed total solar irradiance variation back in time [*Hoyt and Schatten*, 1993; *Lean et al.*, 1995; *Solanki and Fligge*, 1998; 1999; *Lockwood and Stamper*, 1999; *Lean*, 2000] and these reconstructions have been used to evaluate its rôle in the rise of average surface temperatures on *Earth* [*Lean et al.*, 1995; *Lockwood et al.*, 1999b; *Tett et al.*, 1999].

The long-term changes in the Sun also have implications for near-Earth space and for studies of "space climate," i.e., changes in the environment in which modern technological systems, such as spacecraft, operate. The relevant phenomena are caused by energy extracted from the solar wind flow by Earth's magnetosphere, an indicator of which is the level of geomagnetic activity. This has been monitored in a continuous and homogeneous manner since 1868 using antipodal observing stations in southern England and Australia. The level of activity was quantified by *Mayaud* [1972], who developed the *aa* index, using the range of the fluctuations of the horizontal component of the field observed in 3-hour intervals at these stations. The *aa* index was designed to be very well correlated with other planetary indices of geomagnetic activity; for example, the variations in annual means of *aa* and the

[1]Also at Department of Physics and Astronomy, University of Southampton, Southampton, Hampshire, England, United Kingdom.

Paper number 2000JA000115.
0148-0227/01/2000JA9000115$09.00

Ap index are almost identical since the start of the *Ap* data series in 1932 [*Mayaud,* 1972; *Ahluwalia,* 1997]. This is true for both the solar cycle and the longer-term variations. The long-term changes in *aa* are also highly correlated with related phenomena, such as the occurrence of low-latitude aurorae [*Pulkkinen et al.*, 2000]. The *aa* data sequence shows a systematic long-term rise in geomagnetic activity since 1900, a variation which shows some intriguing similarities to the observed rise in global mean surface temperatures [*Cliver et al.*, 1998b]. *Clilverd et al.* [1998] have analyzed in detail a variety of potential causes for this rise in *aa* and eliminated most of them. For example, the long-term drift of the geomagnetic field has caused a systematic shift in the magnetic latitudes of the stations: The Northern Hemisphere station used to generate *aa* has drifted about 4° equatorward in magnetic coordinates since 1868, whereas that in the Southern Hemisphere has drifted about 2° poleward. However, the data show that the activity at both stations has risen in an almost identical manner and thus the rise is not caused by this effect. *Feynman and Crooker* [1978] used the *aa* index to investigate how interplanetary conditions have altered over the past 140 years. *Stamper et al.* [1999] studied interplanetary data from solar cycles 20, 21, and 22 (1963-1996) in a search for the cause of the rise in *aa*. They used the theory of solar wind power extraction by Earth's magnetosphere [*Vasyluinas et al.*, 1982] to show that the largest of several contributions to the long-term change in *aa* was a rise in the magnitude of the interplanetary magnetic field around the Earth. *Lockwood et al.* [1999a] also used this theory to, effectively, separate the effects of nonrecurrent geomagnetic activity (for example, induced by coronal mass ejections) and recurrent geomagnetic activity (due to fast solar wind streams) and thereby developed a procedure to compute the magnitude of the interplanetary field at Earth B_E from the *aa* data.

There is another important indicator of long-term solar change derived from a continuous measurement sequence. This is the solar cycle length L, derived from the sunspot numbers R. *Gleissberg* [1944] noted that there was a long-term (of order 80-100 years) quasi-periodic behavior in R and that this was related to L. He employed a long-timescale secular filter to smooth the values of L, which were determined from the times of adjacent minima and maxima of R. In recent years a variety of other methods to determine L have been developed. For example, *Hoyt and Schatten* [1993] determined from each annual mean of R the fraction of the cycle that had been completed and then measured the delay to the same point in the next cycle. *Fligge et al.* [1999] have employed Morlet wavelets; *Mursula and Ulich* [1998] have used median activity times. *Friis-Christensen and Lassen* [1991] used a method and filter that were similar to Gleissberg's and found that L was anticorrelated with global average temperatures on Earth. The irradiance constructions of *Hoyt and Schatten* [1993] and *Solanki and Fligge* [1998; 1999] depend on L. The sequences of the various estimates of L have some similarities to the observed variations in the *aa* geomagnetic index.

1.2. Coronal Source Flux

The coronal source surface is where the magnetic field becomes approximately radial. It is a roughly spherical surface at a heliocentric distance r of about $2.5R_s$ (where R_s is a solar radius; that is, the photosphere is at $r = 1R_s$). This surface can be considered as the boundary that separates the solar corona from the heliosphere. The magnetic flux threading the coronal source surface is called the open solar flux or the coronal source flux, F_s.

On annual timescales, the interplanetary magnetic field (IMF, the heliospheric field in the ecliptic plane), obeys the

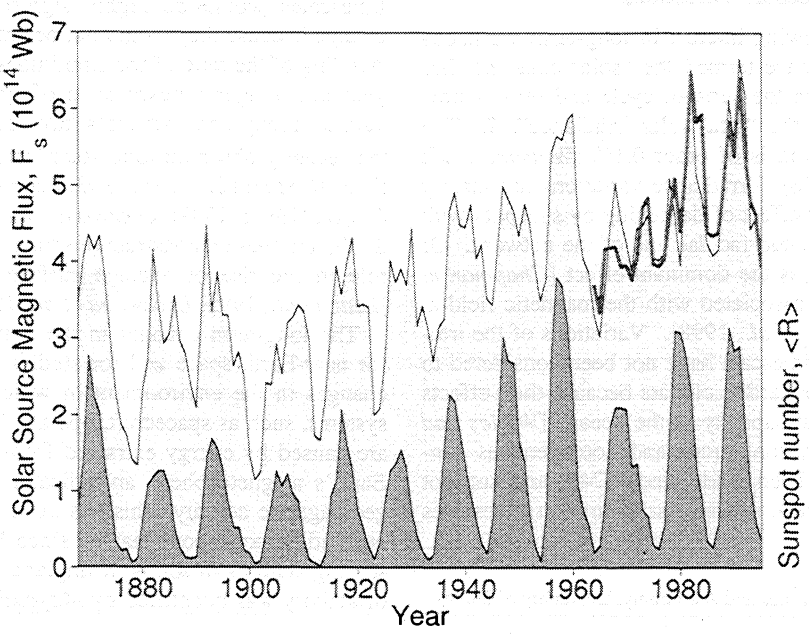

Figure 1. Variation of annual means of the coronal source flux F_s as derived from the *aa* index by the method of *Lockwood et al.* [1999a] (thin line bounding lightly shaded area) and from interplanetary measurements of the radial component of the interplanetary magnetic field near Earth (thick line). The dark shaded area gives the variation of the smoothed sunspot number.

Parker spiral orientation with a garden hose angle γ that is approximately constant at a given r [*Gazis*, 1996; *Stamper et al.*, 1999]. Thus the IMF's radial component near Earth $B_{rE} = B_E \cos(\gamma)$ can also be estimated from the *aa* data using the method of *Lockwood et al.* [1999a]. This has significance for the heliosphere away from the ecliptic plane because the Ulysses spacecraft has shown that sheet, rather than volume, currents dominate, to the extent that latitudinal gradients of the average radial heliospheric field are small [*Balogh et al.*, 1995; *Lockwood et al.*, 1999b]. This means that the radial field B_r, at a heliocentric distance r and at any latitude, is approximately equal to $(r/R_1)^2 B_{rE}$, where $R_1 = 1$ AU. Because, on average, the Parker spiral theory of the heliospheric field applies and because this does not predict significant flux crossing the current sheet at $r < R_1$, the total magnetic flux leaving the Sun and entering the heliosphere (the coronal source flux) is $F_s = (1/2)4\pi R_1^2 \langle |B_{rE}| \rangle$. The method of *Lockwood et al.* [1999a] for deriving annual means of F_s from the *aa* index assumes that both Parker spiral theory and the uniformity of B_r were valid at all times since 1868. The method was developed using data from solar cycles 21 and 22 tested against independent interplanetary measurements from cycle 20 [*Lockwood and Stamper*, 1999]. Their results are shown in figure 1. The lightly shaded area is the coronal source flux estimated from the *aa* index; the thick line is the corresponding value derived from the mean radial field $\langle |B_{rE}| \rangle$ measured by spacecraft near $r = R_1$. The area shaded black gives the variation of smoothed sunspot number for comparison.

Before considering the implications of Figure 1, it is important to note that the coronal source flux has been also been estimated from measurements of the line-of-sight component of the photospheric field (at $r = 1R_s$). In deriving this line-of-sight component of the field from magnetograph data, a latitude-dependent "saturation" correction factor must be applied [*Wang and Sheeley*, 1995]. The radial component is then computed by dividing by a cosine factor (so there is no information from over the solar poles). The open flux is then estimated using a method such as the potential field source surface (PFSS) procedure [*Schatten et al.*, 1969], in which the coronal field is assumed to be current-free between the photospheric surface and the coronal source surface, where the field is assumed to be radial. With an improved latitude-dependent saturation correction factor, *Wang and Sheeley* [1995] were able to match to the radial field seen at Earth (as shown by the thick line in Figure 1) during solar cycles 20 and 21, again using the assumption that B_r is independent of latitude in the heliosphere, as found from the Ulysses observations. In this way, Wang and Sheeley reproduced the upward drift during this period that is revealed in Figure 1. In this paper, we do not consider such estimates from procedures like PFSS, and hereafter, the quantity F_s is defined as the coronal source flux, estimated from the *aa* geomagnetic index using the method of *Lockwood et al.* [1999a].

Figure 1 shows that the average F_s has risen by 34% since 1964 and by 140% since 1900. Such changes could all be in the corona, with little implication for the total flux emerged through the photosphere (which is of order $10F_s$ [*Wang et al.*, 2000b]), nor possibly even for the distribution of that flux over the solar surface. However, the changes in the magnetic field leaving the solar corona, as shown in Figure 1, have marked similarities to some changes seen in the solar photosphere which reflect changes in the magnetic field there and in the subsurface layers. Figure 1 shows that the peak sunspot number at the maximum of the solar cycles has risen in association with F_s, as have the solar-cycle averages of the sunspot number. Both these parameters show very high (and highly significant) correlations with F_s [*Lockwood et al.*, 1999b]. Furthermore, there is a similarly strong correlation between annual means of F_s and the standard deviations of sunspot latitudes [S. Foster and M. Lockwood, Long-Term Changes in the Solar Photosphere Associated with Changes in the Coronal Source Flux, submitted to *Geophys. Res. Lett.*, 2000]. In other words, the photosphere shows more spots, spread over a greater area of its surface, when F_s is higher. Thus the variations in F_s do appear to be a symptom of changes in the surface field. These findings indicate that there have been considerable changes in the processes that generate and distribute photospheric and coronal magnetic fields. The magnetic field at the photosphere includes both open flux (that threads the coronal source surface) and the larger closed flux (that does not) and is generated by dynamo processes in the solar interior, in particular at the base of the convection zone (at around $r = 0.7R_s$). This field emerges in active regions (as bipolar magnetic regions, or BMRs) [*Harvey and Zwaan*, 1993] at a total rate which increases with the number of sunspots. Most of the flux associated with BMRs is annihilated by diffusion toward the neutral line between regions of opposite field polarity, but some is added to the fields associated with the supergranulation network. This field is transported over the solar surface by the differential rotation in the outer Sun, by convection associated with supergranules, and by meridional poleward flow in the surface layers. As the centers of the two polarity regions of a BMR separate in the differential rotation, the magnetic field loops rise through the coronal source surface, and the coronal source flux increases. Open flux accumulates near the poles, where it forms the large coronal holes seen at the subsequent sunspot minimum [*Wang et al.*, 2000a].

Recent theoretical modeling by *Solanki et al.* [2000] predicts a rôle of cycle length in the long-term variation of coronal source flux shown in Figure 1. Solanki et al. used the sunspot number to quantify the rate of flux emergence in BMRs. As discussed above, this flux is either diffusively annihilated at the neutral lines between the opposite field polarity regions or is transferred to the network. The only free input variable in their model is the time constant τ_N for the rate of decay of the open network flux, i.e., the fraction of the network field continuing through the coronal source surface into interplanetary space. If τ_N is sufficiently large, the cycle length L will influence the amount of residual open network flux because shorter cycles will mean that this decay does not progress as far as during longer cycles. Solanki et al. find that a value for τ_N of 4 years allows them to closely match the variation of the coronal source flux F_s as derived by *Lockwood et al.* [1999a]. Similarly, Foster and Lockwood [Long-Term Changes in the Solar Photosphere Associated with Changes in the Coronal Source Flux, submitted to *Geophys. Res. Lett.*, 2000] derive a value for τ_N of 3.8 years and these estimates are consistent with the numerical modelling of BMR evolution by *Wang et al.*, [2000a]. Thus the modelling by *Solanki et al.* [2000] strongly implies that the cycle length L is a crucial parameter in determining long-term variability of the solar and heliospheric magnetic fields. In order to determine the variation of the solar-cycle length, *Friis-Christensen and Las-*

sen [1991] employed the intervals between successive maxima and minima. However, this produces a noisy variation and, following *Gleissberg* [1944], Friis-Christensen and Lassen passed the data though a complex and long timescale smoothing filter. In section 2 of the present paper, we employ a different procedure to evaluate cycle length and derive a similar, yet significantly different, variation of L.

1.3. Cosmic Rays

Other studies have used a variety of paleomagnetic and paleoclimatic evidence for long-term change in the Sun and in the heliosphere. For example, the heliospheric field has a great rôle in shielding the Earth from galactic cosmic rays [e.g., *Moraal,* 1993, *Cane et al.,* 1999]. The Beryllium isotope ^{10}Be is formed as a spallation product in the upper atmosphere when galactic cosmic rays impact oxygen and nitrogen nuclei. Via precipitation, this isotope is deposited in ice sheets, and its abundance in ice cores allows studies of long-term change on timescales of 500 years and greater [*Beer et al.*, 1998]. Cosmic rays have been observed indirectly using neutron monitors (for which corrections must be made for lead in the ground, solar neutrons, and air mass), by ground-based ionization chambers and by Geiger counters carried on daily balloon flights. These measurements have been made for up to six solar cycles at a number of locations. The amplitude of the solar cycle variation varies with location, as expected because of the additional shielding effect of the Earth's magnetic field depends on geomagnetic latitude, but the different monitoring instruments have detected solar cycle variations that are very similar in form. In section 3 we evaluate the relationship between F_s and galactic cosmic rays at different energies, and in section 4 we investigate the implications of the ^{10}Be record for the longer-term variation of the coronal source flux.

1.4. Oscillations in the rotation rate of the solar interior

The interior solar motions associated with the solar dynamo have been probed using the helioseismology technique. In particular, they have revealed an outer meridional circulation and the tachocline, a shear in the longitudinal flow around $r = 0.7R_s$ separating the core from the differentially-rotating convection layer. Recent observations using the Solar and Heliospheric Observatory (SoHO) spacecraft and the Global Oscillation Network Group (GONG) ground-based network have revealed a 1.3-year oscillation in which the equatorial super-rotating outer layers (at $r > 0.7R_s$) speed up a little, while the core correspondingly slows down (at $r < 0.7$ R_s), and vice versa [*Howe et al.*, 2000]. In section 5 we look for signatures of this periodicity in the heliosphere and study its long-term variations using the *aa* geomagnetic data sequence.

1.5. Implications

In the final section of this paper we consider the implications of the long-term changes in the solar and heliospheric fields that are revealed by the *aa* geomagnetic index and the abundance of isotopes generated by cosmic ray bombardment. These include the potential impact of cosmic rays on cloud cover. In addition, we contrast the techniques used to reconstruct the variation of solar irradiance over the past 140 years. The method of *Lean et al.* [1995] is based on sunspot numbers with a superposed long-term drift, the amplitude of which is inferred from comparison of Sun-like stars. The method of *Solanki and Fligge* [1998; 1999] uses sunspot numbers to account for the contribution of active regions and solar cycle length L to estimate the contribution of the quiet Sun. In comparison, the extrapolation by *Lockwood and Stamper* [1999] uses a simple, single correlation between the irradiance and the coronal source flux F_s estimated from the

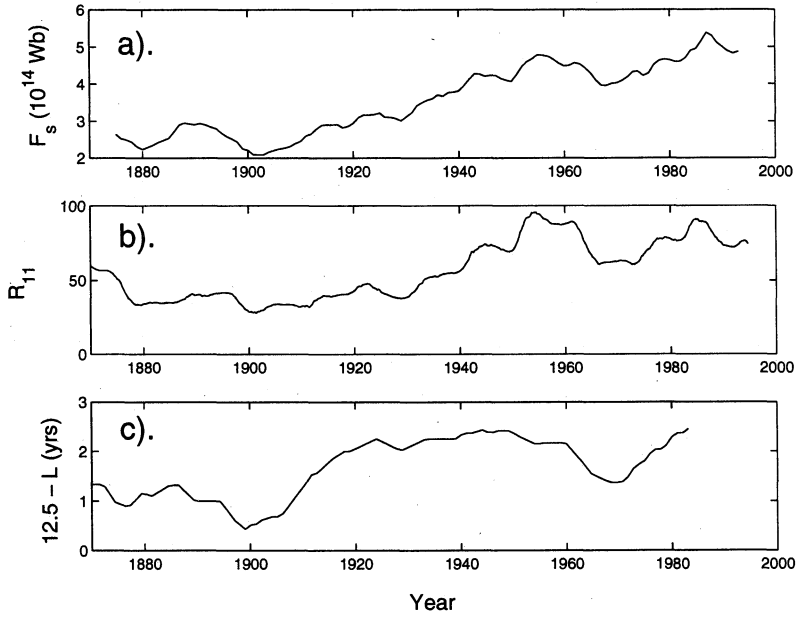

Figure 2. (a) The 11-year running means of the coronal source flux F_s, derived from the *aa* index of geomagnetic activity (see Figure 1); (b) 11-year running means of the sunspot number, R_{11}; (c) $(12.5 - L)$, where L is the length of the solar cycle in years and is derived from autocorrelation functions of sunspot number R, such as the one shown in Figure 3.

aa index. Recently, *Wang et al.* [2000b] have pointed out that the open flux is a very small and variable fraction of the photospheric field and thus its use as a proxy for solar irradiance does not have a firm theoretical foundation. Nevertheless, it is striking how similar the reconstructions of *Lean et al.* [1995], *Solanki and Fligge* [1998, 1999], and *Lockwood and Stamper* [1999] are, especially considering that they employed entirely different proxies. In section 6.3 we discuss how the modeling by *Solanki et al* [2000] may offer a potential explanation of why L and F_s can be used as proxies for irradiance.

2. Coronal Source Flux, Sunspot Number, and Solar Cycle Length

Figure 2 shows three indicators revealing solar change over the past 150 years. Figure 2a shows 11-year running means of the coronal source flux F_s derived from the *aa* index of geomagnetic activity (see Figure 1). Figure 2b shows 11-year running means of the sunspot number, R_{11}. As discussed by *Lockwood et al.* [1999b], the correlation between the two is high and has very high statistical significance. The model of *Solanki et al.* [2000] predicts such an association because the rate with which flux emerges in BMRs is proportional to $(1+A_f/A_s)R$ where A_f is the area covered by faculae and A_s is the area covered by sunspots. The factor in parentheses allows for the fact that not all the newly emerged flux in active regions takes the form of sunspots. Because the ratio (A_f/A_s) can be characterized as a simple function of R [*Chapman et al.*, 1997; *Fligge et al.*, 1998], the total flux emergence rate in active regions is also a function of R. This is one of the two major factors determining the variation of F_s. However, as discussed in section 1.2, the solar cycle length L will also have an influence via the loss of open network flux. Figure 2c shows $(12.5 - L)$, where L is the length of the solar cycle in years. (Note that the value of 12.5 is simply an arbitrary and convenient reference value and has no significance.)

Figure 2c is derived from the autocorrelation functions of the monthly means of the sunspot number R, such as the one shown in Figure 3. The R data sequence for 20-year intervals is correlated with the data from 20-year periods that have been lagged by between -20 years and zero. In order to determine the cycle length L the peak of the autocorrelation function near a lag of −11 years is fitted with a polynomial (otherwise, small fluctuations near the peak can shift the value of L derived). Error bars can be placed on this estimate using the Fischer-Z test by looking for a difference in the correlation coefficient that is significant at the 95% level: These errors are typically ±2 months. In the case shown in Figure 3 (for unlagged data from 1980 to 2000) the peak is at a lag of −10.26 years, giving a cycle length $L = 10.26 \pm 0.20$ years. The values of L are taken to apply to the interval between the midpoint of the original unlagged data series and a time L earlier. For figure 3, this interval is between $(1990 - L) = 1979.74$ and 1990, which is centered on $(1990 - L/2) = 1984.87$.

The L variation in Figure 2c has some similarities to those derived by other methods [e.g., *Friis-Christensen and Lassen*, 1991; *Hoyt and Schatten*, 1993] in that $\{12.5 - L\}$ (where L is expressed in years) shows a rise since 1900, interrupted by a fall between about 1950 and 1970. However, the rise is not as uniform as, for example, that derived by Friis-Christensen and Lassen. In particular, in Figure 2, $\{12.5 - L\}$ rises more rapidly in the interval 1900-1920 and then stays relatively constant for 1920-1960. These differences arise from the method of derivation. Which of the several methods to determine cy-

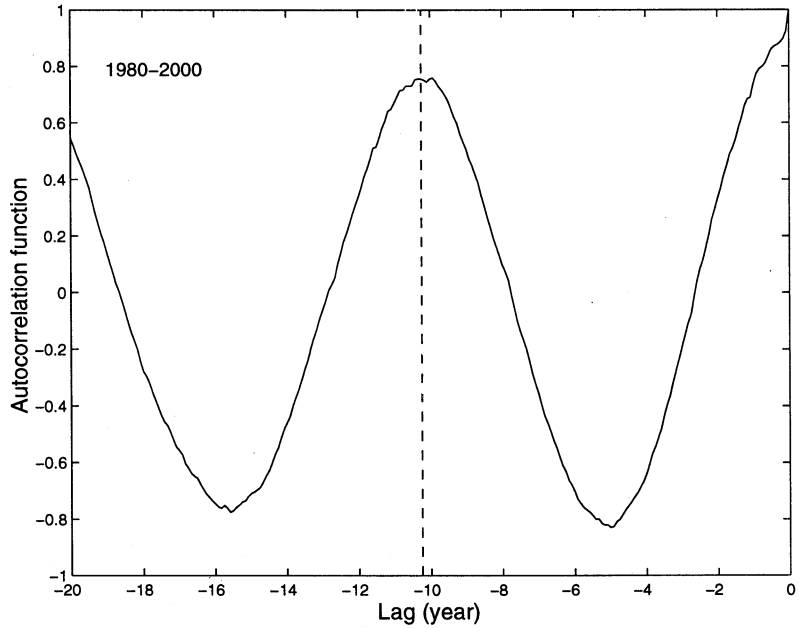

Figure 3. Determination of solar cycle length L, using the autocorrelation function of the monthly sunspot number R. In this example, the R data sequence for 1980-2000 is correlated with the data from 20-year periods from (1980+lag) to (2000+lag) for lag values between -20 years and zero. The peak near -11 years is fitted with a polynomial to determine the lag of peak autocorrelation (dashed line). In this case, the peak is at a lag of −10.26 years, giving a cycle length $L = 10.26$ years. This value of L is taken to apply to the interval between (1990 - L) and 1990, which is centered on (1990 - $L/2$) = 1984.87.

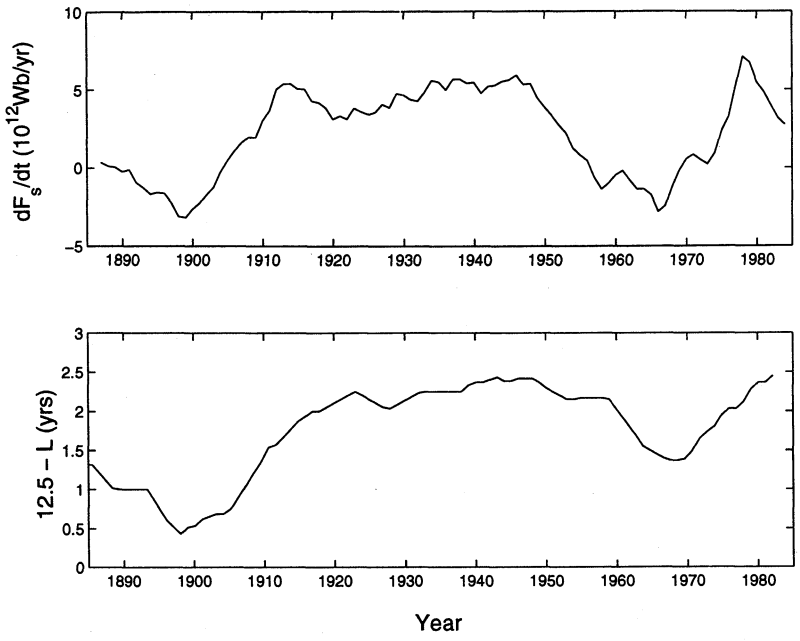

Figure 4. (top) Rate of increase in the coronal source flux (dF_s/dt). Running means over 20-year intervals are presented. (bottom) The variable $\{12.5 - L\}$, where L is the length of the solar cycle, in years, derived from the autocorrelation function of sunspot number R using 20-year intervals (as in Figure 2).

cle length is the most reliable is not clear; however, the method we employ here does have the advantage that it uses data from all phases of the solar cycle and is not so highly dependent on the precise timings of the sunspot minima and maxima.

To investigate the relationship to F_s Figure 4 compares the variation of $\{12.5 - L\}$ (bottom graph) with the rate of increase in the coronal source flux, dF_s/dt (top graph). Because 20-year intervals were used to evaluate L, running means of dF_s/dt over 20-year intervals are presented. A clear similarity is evident. In particular, when L is a maximum (minima in $\{12.5 - L\}$), dF_s/dt is negative. Some lag is present in Figure 4, but considering that these are smoothed data sets, this is probably not significant. This relationship between the cycle

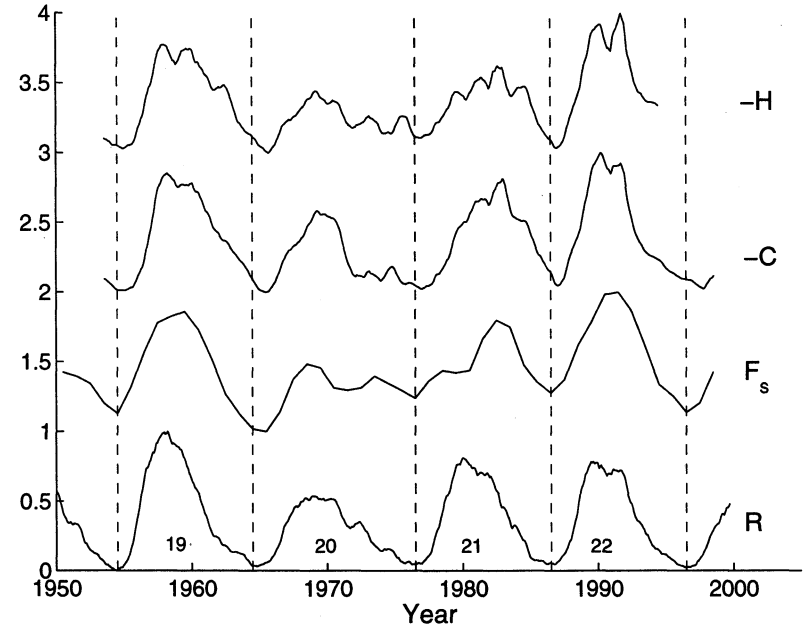

Figure 5. Normalized variations for 1950-2000: The sunspot number R, the coronal source flux F_s derived from the *aa* index, the Climax cosmic ray neutron monitor counts C (inverted to stress anticorrelation with F_s in particular), and the Huancayo/Hawaii cosmic ray neutron monitor counts H (also inverted). The vertical dashed lines show the times of sunspot minima, and the sunspots cycles are numbered. All variations have been normalized to unity amplitude.

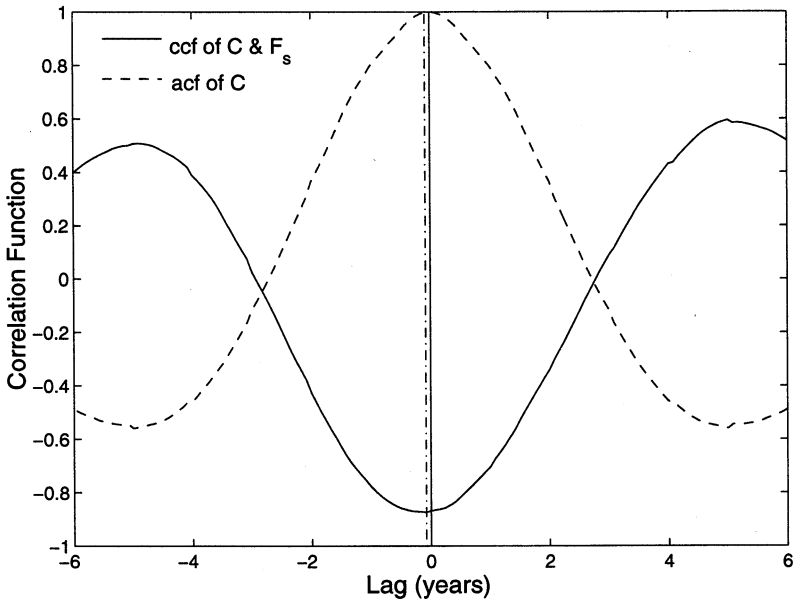

Figure 6. Autocorrelation coefficient of the cosmic ray counts observed at Climax C, as a function of lag (dashed line) and the cross correlation of C with the coronal source flux F_s (solid line). A positive lag is defined as C leading F_s and peak cross correlation is shown by the dash-dotted line at a lag of -1 month. Both F_s and C values are annual means for 1953-1999.

length L and dF_s/dt was inherent in the modeling of the F_s data sequence by *Solanki et al.* [2000]. Thus Figure 4 confirms their conclusion that cycle length L and the rate of flux emergence (related to sunspot number R) in active regions are both key elements in the long-term accumulation of open solar flux (when average sunspot numbers R_{11} are high and/or when cycle lengths L are low) and its decay (when R_{11} is low and/or when cycle lengths L are large).

3. Coronal Source Flux and Cosmic Ray Fluxes

The Earth is shielded from galactic cosmic rays by the heliosphere. A number of processes are active, although their relative importance is still a matter of debate. However, the heliospheric field is the key component of this shield [*Moraal et al.,* 1993; *Potgieter,* 1995], such that the cosmic ray fluxes seen at Earth are very highly anticorrelated with the interplanetary magnetic field [*Cane et al.*, 1999]. This has been seen at a variety of heliocentric distances, with a delay consistent with the propagation of changes in the heliospheric field from the coronal source surface to the point of observation at the speed of the solar wind flow [e.g., *Potgieter,* 1995]. This delay would have a maximum of order one year for an observation point close to the heliospheric termination shock [*Cummings, et al.*, 1994]. Because of their large gyroradii and complex paths, cosmic rays sample much greater regions of the heliosphere than do near-Earth satellites. Thus we would expect a strong anticorrelation of the coronal source flux F_s with cosmic ray fluxes. Galactic cosmic rays have been detected by neutron monitors continuously since 1953, and the normalized variations of the corrected count rate from Climax C and Huancayo/Hawaii H are shown in Figure 5. Negative values are plotted to stress the anticorrelation with other parameters shown, namely the sunspot number R and the coronal source flux estimate F_s. All variations shown have been normalized to a range of unity over the period shown.

Figure 6 shows the autocorrelation coefficient of the annual means of the cosmic ray counts observed at Climax C in 1953-1999, as a function of lag (dashed line) and the cross correlation of C with the coronal source flux F_s (solid line). A positive lag is defined as C leading F_s and peak cross correlation is shown by the dash-dotted line at a lag of -1 month. The peak anticorrelation is –0.874, and the probability of this being a chance result is negligible. The scatter plot of the annual means of C and F_s at this lag is shown in Figure 7. The line is the best-fit, least squares linear regression, given by $C/10^3 = 5.22 - 0.278F_s$. Figures 8 and 9 show, respectively, the corresponding correlation functions and the scatterplot of the cosmic ray counts observed at Huancayo/Hawaii H. These two stations provide a homogeneous data sequence, the data series being continued at Haleakala, Hawaii, after monitoring ceased at Huancayo, Peru, in 1993. The peak anticorrelation is at -6 months and is –0.897. The best-fit least squares linear regression is $H/10^3 = 1.866 - 0.345F_s$.

The Climax and Huancayo/Hawaii data differ in a number of ways. Most importantly, the geomagnetic latitude of the two stations is different, which means that they have geomagnetic rigidity cutoffs of 3 and 13 GV, respectively. This means that the products of the primary cosmic ray particles of energy exceeding about 3 GeV are seen at Climax, whereas only those from primaries exceeding about 13 GeV are detected at Huancayo/Hawaii. Using the high correlations and the corresponding least squares regression fits between C and F_s (as shown in Figures 6 and 7) and between H and F_s (as shown in Figures 8 and 9), along with the data sequence of F_s shown in Figure 1, the cosmic ray fluxes can be extrapolated back to 1868. The results are shown in Figure 10, the solid line being for the Climax data (>3 GeV), and the dashed line being for the Huancayo/Hawaii data (>13 GeV). In both cases, the count rates have been normalized to the average value seen during solar cycle 21. The plot indicates that the average fluxes of cosmic rays above 3 GeV were approxi-

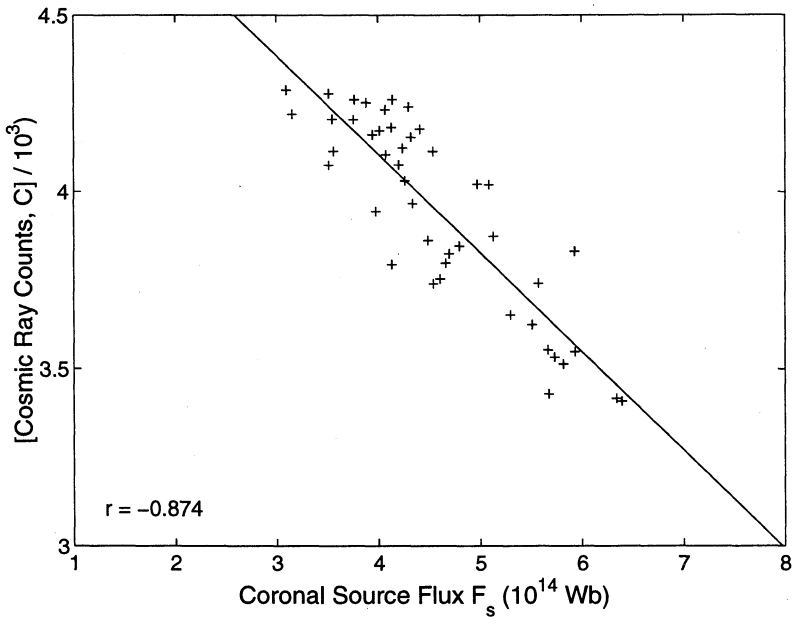

Figure 7. Scatter diagram of annual means of F_s and C, the latter lagged by 1 month, which gives a peak anti-correlation of -0.874. Data are for 1953-1999 (inclusive). The best-fit, least squares linear regression line shown is $C/10^3 = 5.22\text{-}0.278F_s$.

mately 15% higher in 1900 than they are now, whereas the fluxes above 13 GeV were higher by about 4%.

Also shown in Figure 10 is the variation of cosmic ray fluxes since 1937 deduced from a collection of ionization chambers at high latitudes, and near sea level [*Ahluwalia*, 1997] (solid line joining dots). The sites employed are Cheltenham (1937-1956), Fredericksberg (1956-1972) and Yakutsk (1954-1994). The geomagnetic rigidity cutoffs at these stations are 2.2 GV, 2.2 GV and 1.7 GV, respectively, but fluxes are limited by the higher atmospheric cutoff of about 4 GV in each case. The muons detected relate to somewhat higher energy primary cosmic rays than for the neutron monitors discussed above, and the median rigidity observed by these ionization chambers is 67 GV. Figure 10 shows that the variation is similar to that obtained from the correlation of F_s and cosmic rays at >13 GV. To provide a single sequence,

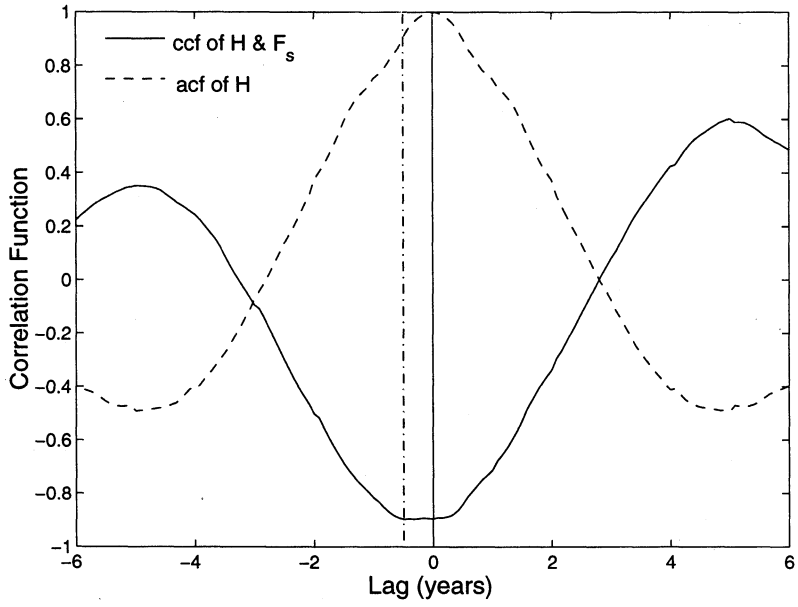

Figure 8. The autocorrelation coefficient of the cosmic ray counts observed at Huancayo/Hawaii H, as a function of lag (dashed line) and the cross correlation of H with the coronal source flux F_s (solid line). A positive lag is defined as H leading F_s, and peak cross-correlation is shown by the dash-dotted line at a lag of -6 months. Both F_s and H values are annual means for 1953-1995.

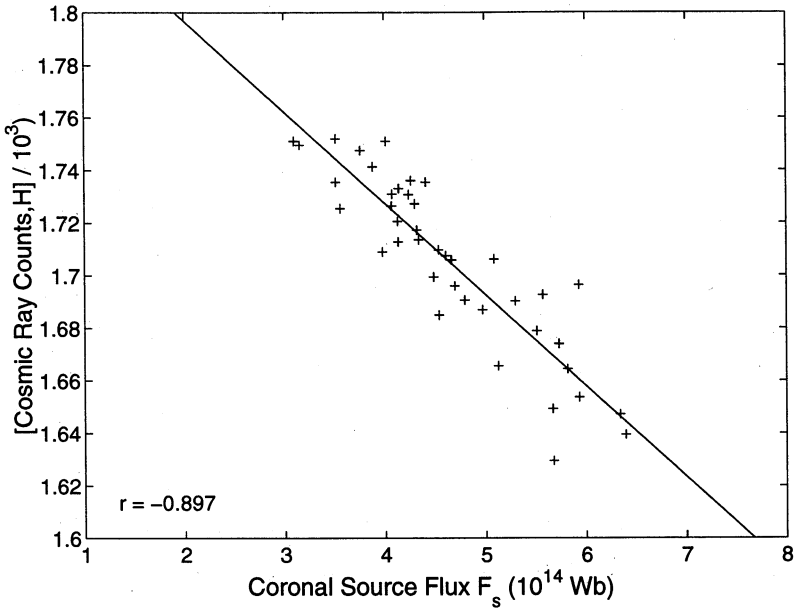

Figure 9. Scatter diagram of annual means of F_s and H, the latter lagged by 6 months, that gives a peak anti-correlation of -0.897. Data are for 1953-1995 (inclusive). The best-fit least squares linear regression is $H/10^3 = 1.866\text{-}0.345F_s$.

considerable intercalibration factors have been applied to the data from these ionization chambers [*Ahluwalia*, 1997], and thus long-term drifts may not be accurately represented. Of particular concern is that so little overlap in data (1954-1956) exists for full annual means from Cheltenham and Yakutsk. Nevertheless, there are strong similarities for the variation for >13 GV predicted from the F_s values (the correlation coefficient is 0.811). In terms of the percentage change, both the solar cycle variations and the longer-term drifts seen by the ionization chambers are smaller (indicating that they are responding to higher-energy primary cosmic rays that are less influenced by the heliospheric shield). Analysis of global car-

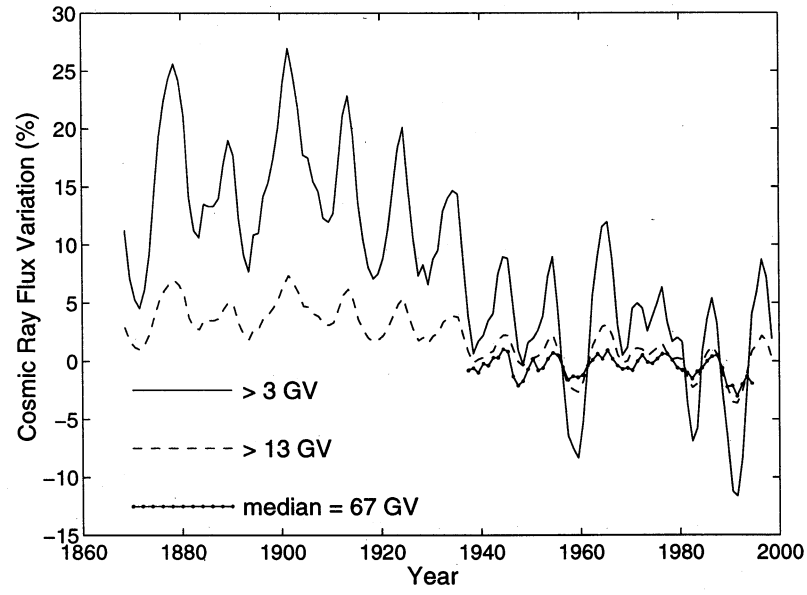

Figure 10. Inferred variation of cosmic ray fluxes since 1868. The solid line is an extrapolation using F_s, based on the correlation shown in Figure 6 with counts C by the cosmic ray neutron detector at Climax (geomagnetic rigidity cutoff 3 GV). The dashed line is based on the correlation, shown in Figure 11 with the counts H by the cosmic ray neutron detectors at Huancayo/Hawaii (geomagnetic rigidity cutoff 13 GV). The line joining dots is the variation deduced from three ionization chambers by *Ahluwalia* [1997], which respond to a median rigidity of 67 GV. In all cases, the variation is relative to the average value seen during solar cycle 21.

bon-14 production [*Struiver and Quay*, 1980] implies that the combined ionisation chamber data may underestimate a slight downward drift in the cosmic ray fluxes in the period 1937-1970 [*O'Brien*, 1979].

The correlations shown in Figures 6-10 have implications for understanding cosmic ray retardation in the heliosphere. Several authors have noted that there appears to be a 22-year cycle in cosmic ray behavior. This is, to some extent, expected theoretically because the heliospheric field changes polarity shortly after the peak of each sunspot cycle. When the heliospheric field is directed outward in the northern hemisphere (when we define the polarity A to be positive) positively charged cosmic rays will drift from polar regions to the equatorial current sheet of the heliosphere, whilst negatively charged cosmic rays drift in along the current sheet and then to polar latitudes. When the heliospheric field polarity A is less than zero, these directions are reversed. The current sheet becomes increasingly warped at higher sunspot numbers because of the increase in the tilt angle. As a result of the dependence of the drifts on the field polarity A, the flux of positive cosmic ray particles should be relatively insensitive to variations in the tilt angle when $A > 0$ but would be a stronger function of tilt angle for $A < 0$. These drift effects are expected to be most important in the years around sunspot minimum, when the effect of outward propagating diffusive barriers is reduced because of the lower occurrence of the coronal mass ejections (CMEs), high-speed flows, and shocks that merge into such barriers.

The data since 1953 show that the recovery of cosmic ray fluxes does take longer as sunspot minimum is approached in even-numbered sunspot cycles (when $A < 0$) than in odd-numbered cycles (when $A > 0$) [e.g., *Ahluwalia and Wilson*, 1996; *Usoskin et al.*, 1998]. However, the decay of fluxes in the rising phase of cycle 21 was also delayed, and Figure 5 shows that the lag in both the fall and the rise in the cosmic ray fluxes was particularly large for sunspot cycle 21, but this additional lag is not nearly so clear for cycle 19. The same behavior is apparent for F_s. *Cane et al.* [1999] have found that the difference of the tilt angle dependence for $A < 0$ and $A > 0$ is restricted to a small range of tilt angles around solar minimum and conclude that much of the good correlation between the tilt angle and the cosmic ray intensity at other parts of the solar cycle is due to the fact that the tilt angle and the heliospheric field strength vary roughly in concert during the solar cycle. Thus they conclude that, apart from a short period around sunspot minimum, there is no difference between the odd and even cycles and that it is the heliospheric field that controls the cosmic ray fluxes. This is consistent with the good correlations shown in Figures 7 and 9, which contain data for both polarities of the heliospheric field. The correlation coefficients of $c = -0.874$ and -0.879, show that F_s, on its own, explains $c^2 = 76.4\%$ and 77.3% of the variation in the cosmic ray fluxes seen at Climax and Huancayo/Hawaii, respectively. This sets an upper limit of about 23% for the combined effect of polarity-dependent drifts, as discussed above, and of the solar wind speed [*Sabbah*, 2000]. The correlation coefficient of $c = 0.811$ means that $c^2 = 66\%$ of the variation seen by the ionisation chambers (Figure 10) is explained by F_s alone.

It is therefore instructive to look at the lag between cosmic ray fluxes, F_s, and R in other sunspot cycles. Figure 1 shows that for cycles 11-22, the lag between F_s and R is the greatest for cycle 21. This implies that cycle 21 was unusual in that the peak in F_s was exceptionally delayed after the peak of the sunspot cycle [see also *Wang et al.*, 2000b]. There is also (smaller) delay for cycle 19, but again, this is seen in both the

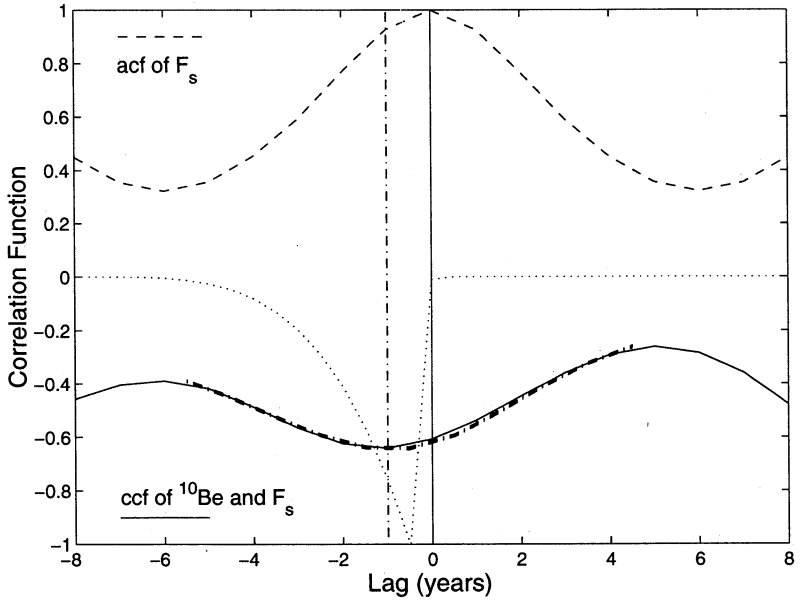

Figure 11. Autocorrelation coefficient of the coronal source flux F_s as a function of lag (dashed line) and the cross correlation of F_s with the ^{10}Be isotope abundance (solid line). A positive lag is defined as ^{10}Be abundance leading F_s, and peak cross-correlation is shown by the dash-dotted line at a lag of T =-1 year. Both the F_s and the ^{10}Be abundance values are annual means for 1868-1985. The dotted line gives the best estimate of the response function, with a peak at T = -0.5 year and a $(-T)^4$ decay over 5 years. This gives a best fit to the cross-correlation function which is shown by the thick dash-dotted line.

cosmic ray fluxes and in F_s (see Figure 5). Thus the limited sample of four solar cycles of homogenous cosmic ray data, including what appears to be an exceptional cycle (cycle 21), may have led us to overestimate the effect of polarity-dependent drifts. The delay in the minimum cosmic ray flux after sunspot maximum is indeed greater for the two odd-numbered cycles monitored (cycles 19 and 21) [*Ahluwalia and Wilson*, 1996; *Usoskin et al.*, 1998], but this appears to be caused by a corresponding delay in the peak of F_s. Figure 1 shows that large lag (>1 year) between the peaks in R and F_s occurs for cycles 14, 15, 19, and 21. Thus there is a tendency for this lag to be present in odd-numbered cycles, but there are exceptions to this (cycle 14 shows a lag, but cycles 11, 13, and 17 do not). The reason for this tendency is not known. However, the values of F_s are annual means and thus often this lag is not well resolved. The model of *Solanki et al.* [2000] does predict peaks in the open flux that are delayed, compared to the F_s peaks shown in Figure 1, for cycles 11, 13, and 17. This implies that the asymmetry between odd and even cycles is present in the flux emergence rate and must also be present, in a subtle manner, in the sunspot numbers, from which Solanki et al. computed emergence rate.

4. Longer-Term Variations of Coronal Source Flux and Cosmic Rays

The high anticorrelations between cosmic rays and the coronal source flux discussed in section 3 imply that there should be a strong anti-correlation between F_s and the abundance of the ^{10}Be isotope in ice cores, produced by cosmic ray bombardment. Figure 11 shows the autocorrelation coefficient of the coronal source flux F_s as a function of lag (dashed line) and the cross-correlation of F_s with the ^{10}Be isotope abundance (derived from the Dye-3 Greenland ice core [*Beer et al.*, 1998]). Both the F_s and the ^{10}Be abundance values are annual means for 1868-1985. A positive lag is defined as ^{10}Be abundance leading F_s , and peak anticorrelation is shown by the thin dash-dotted line at a lag of $T = -1$ year. The ^{10}Be isotope is formed as a spallation product in the upper atmosphere when galactic cosmic rays impact oxygen and nitrogen nuclei. This isotope is deposited in ice sheets by precipitation over a subsequent extended period. Thus a lag is expected. The Wiener-Lee theorem states that the cross-correlation function (ccf) between the output and input of a system is equal to the convolution of the autocorrelation function (acf) of the input with the response function of the system. The dotted line in Figure 11 gives our best estimate of the response function of the ^{10}Be deposition, with a peak at $T = -0.5$ year and a $(-T)^4$ decay over 5 years. The thick dash-dotted line is the best-fit convolution of this response function with the acf of the input (i.e., of F_s).

Figure 12 shows the scatterplot of coronal source flux F_s and the concentration of the ^{10}Be isotope for 1868-1985, with the lag of 1 year which gives the peak correlation coefficient of –0.64. The line is the best-fit, least squares linear regression, which is [^{10}Be in atoms g^{-1}] $= 10^4$ (1.323 – 0.133F_s). Figure 13 compares the extrapolated variation of cosmic ray flux (at >13 GV) since 1868 (solid line, as shown in Figure 10), with the ^{10}Be isotope record, scaled using the correlations shown in Figures 7 and 12. In both cases, the variation is relative to the average value seen during solar cycle 21. The long-term trend is reproduced in both data sets although the solar cycle variations are not identical. This is consistent with the results of *Fligge et al.* [1999], who found that cycle lengths derived from the ^{10}Be isotope data were not fully reliable, presumably because the extended response function is

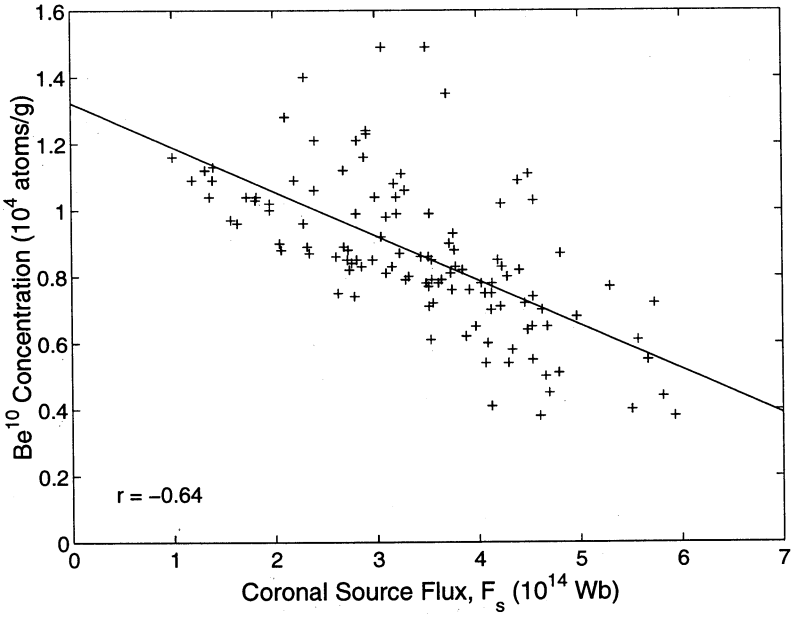

Figure 12. Scatterplot of coronal source flux F_s and the concentration of the isotope ^{10}Be deposited in the Dye-3 Greenland ice core. Data are annual means for 1868-1985. The ^{10}Be data are 10-year running means that have been lagged by 1 year to give the peak correlation coefficient of -0.64. The best-fit least squares linear regression is [^{10}Be atoms g^{-1}]/10^4 = 1.323 – 0.133F_s.

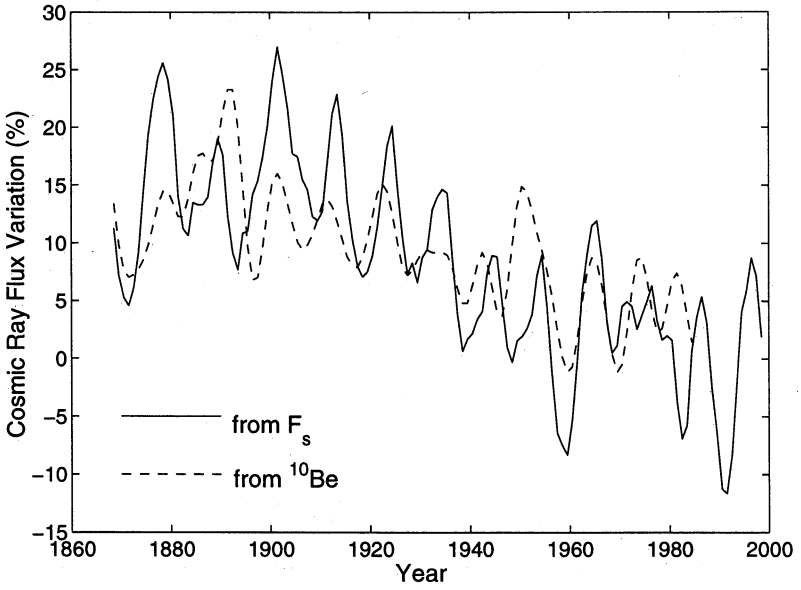

Figure 13. Variation of cosmic ray flux since 1868. The solid line is an extrapolation using F_s, based on the correlation with fluxes at >3 GeV shown in Figure 7. The dashed line is based on the ^{10}Be isotope record, scaled using the correlation shown in Figures 7 and 12. In both cases, the variation is relative to the average value seen during solar cycle 21.

variable (due to a spread of isotope deposition timescales caused by climate variability). This leads to much of the scatter in Figure 12.

We can use the full data sequence of the ^{10}Be isotope data, along with the regression shown in Figure 12, to estimate the coronal source flux back to near 1400. This extrapolation is shown in figure 14: These data are 11-year running means in which the solar cycles have been smoothed out. Comparison with Figure 2 shows that the drift in average F_s since 1868 is very well reproduced. The plot shows that variations in F_s seen since 1900 are similar in amplitude to those seen at prior times. Particularly rapid changes are seen late in the "Maunder minimum" of sunspot activity (roughly 1645-1715) [e.g., *Cliver et al.*, 1998b], with F_s falling to its lowest value of 1.2 $\times 10^{14}$ Wb near 1700, toward the end of this period. This corresponds to a value of about one quarter of present-day values. That F_s should reach a minimum at the end, rather than in the middle, of the Maunder minimum is consistent with the model of *Solanki et al.* [2000]. This is because the F_s would decay throughout the minimum because the flux emergence rate in active regions was exceptionally low.

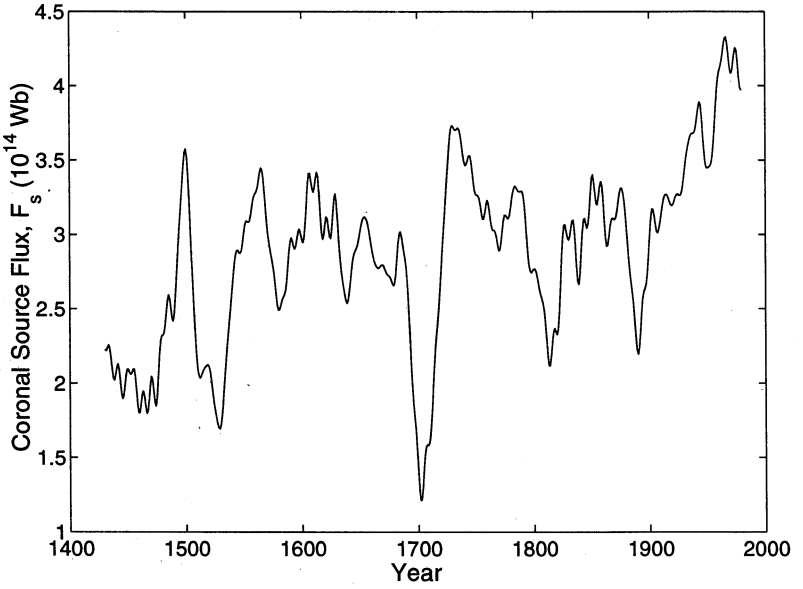

Figure 14. Extrapolated coronal source flux F_s from the ^{10}Be isotope data, using the regression analysis shown in Figure 12.

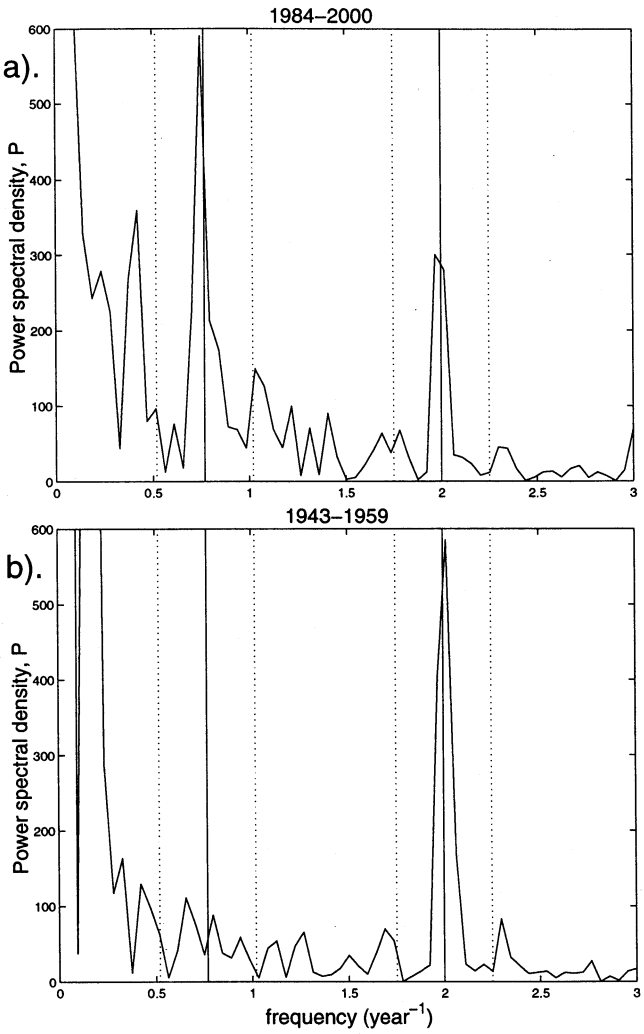

Figure 15. Power spectrum of variations in the *aa* index for (a) 1984-2000 and (b) 1943-1959. The solid vertical lines show frequencies of (1/0.5) and (1/1.3) yr $^{-1}$. The vertical dotted lines mark bands of width 0.5 yr $^{-1}$ about these two frequencies, used to estimate the background variation level.

Other studies also confirm the drift of F_s discussed here. For example, *Bonino et al.* [1995] use the Titanium isotope ^{44}Ti, produced in meteorites by galactic cosmic rays and having a half life of about 96 years, to deduce a rise in the heliospheric field throughout the 20th century: the rise found is four times larger than would be deduced from sunspot numbers using the anticorrelation with cosmic ray fluxes seen since 1953.

5. Variations at a Period of 1.3 Years

Howe et al. [2000] have recently used the helioseismology technique to define a 1.3-year period oscillation in the internal motions of the Sun, close to the base of the convection zone. The data were recorded between May 1995 and November 1999 by the Michelson Doppler Imager (MDI) instrument on the Solar and Heliospheric Observatory (SoHO) spacecraft and by the Global Oscillation Network Group (GONG) network of 6 ground-based observatories. The identification of a 1.3-year oscillation is tentative because the data extend over just four of such cycles. Nonetheless, the periodicity is seen as a variation in the equatorial speed of rotation in the convective zone which is in antiphase with an oscillation in the corresponding speed of rotation of the core, on the other side of the tachocline. Oscillations at 1.3 years have been reported previously in the heliosphere. *Richardson et al.* [1994] and *Gazis et al.* [1995] found variations at this period in the solar wind speeds at heliocentric distances between 0.72 and 60 AU. They deduced a solar origin for these variations. In addition, fluctuations at this period have been seen in the southward component of the IMF [*Szabo et al.*, 1995] and in geomagnetic [*Paularena et al.*, 1995] and auroral phenomena [*Silverman and Shapiro*, 1983].

Power spectral analysis of the sunspot numbers does not reveal a consistent variation at this period. Peaks in the power spectral density form at nearby frequencies, broaden, and move through the 1.3-year period, and almost never is there a persistent, clear peak at 1.3 years. Furthermore, the sunspot data for 1995-2000, the interval for which the 1.3-year oscillation was detected by SoHO and GONG, reveal very low spectral power at 1.3 years. However, the situation is different when we look at terrestrial parameters that depend on the open solar flux. In particular, the near-Earth IMF and the *aa* index both reveal a clear peak in the power spectrum at 1.3 years for 1995-2000. This is illustrated by Figure 15, which shows the power spectra of the *aa* index for data taken in the intervals 1984-2000 (Figure 15a) and 1943-1959 (Figure 15b). Periods of 0.5 and 1.3 years are marked with solid vertical lines. Note that the dominant power is at very low frequencies (~1/11 yr^{-1}) because the solar cycle is the largest variation. In both cases, there is a clear line at a period of 0.5 years. This is expected because of the "Russell-McPherron effect", which is due to the fact that the energy coupling between the solar wind and the magnetosphere to peaks at each equinox because of the dipole tilt of Earth's magnetic field is then most favorable [*Russell and McPherron*, 1973]. In addition, there may be a small semiannual variation introduced into the *aa* index by the seasonal conductivity variations at each site [*Stamper et al.*, 1999]. For the more recent of the two intervals shown in Figure 15, a strong spectral peak is also seen at 1.3 years, a peak which is entirely absent in the data from the earlier interval. Thus the 1.3-year periodicity was not always as strong as it has been in recent years. Study of the full series of spectra such as those in Figure 15 shows that the peak grows and falls in amplitude at the period of 1.3 years, rather than evolving from adjacent periods.

To investigate the past history of these periodicities, we need to quantify the strength of the variation at a given period. In many cases, the spectral lines are not as clear as in Figure 15, often because equally strong fluctuations are present at other nearby frequencies. We here normalize the strength of the frequency at the period of interest by also looking at the power at nearby frequencies. If we are interested in a period of X, we here derive a normalized power $P^*(X^{-1})$ which we define to be (P_l/P_b) where P_l is the power spectral density in the spectral line, integrated over a narrow frequency range of 0.2 yr $^{-1}$ about the frequency $(1/X)$ (i.e., in the frequency range $(1/X) - 0.1$ to $(1/X) + 0.1$ yr $^{-1}$). P_b is the average power in the background surrounding the line which we take to be in the rest of the frequency band of 0.5 yr $^{-1}$ about $(1/X)$ (i.e., in the frequency ranges $\{(1/X) - 0.25\}$ to $\{(1/X) - 0.1\}$ and $\{(1/X) + 0.1\}$ to $\{(1/X) + 0.25\}$ yr $^{-1}$). Thus

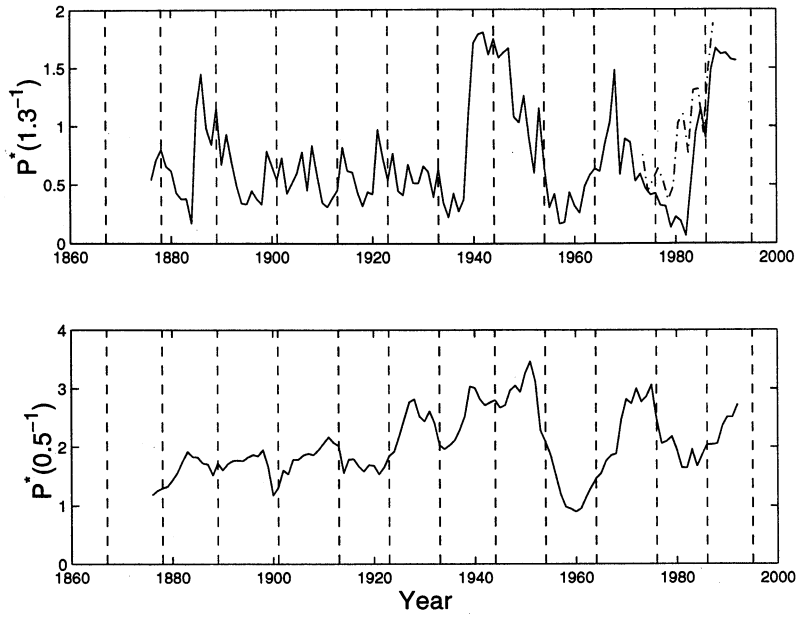

Figure 16. (top) Normalized power spectral density at a period of 1.3 years, $P^*(1.3^{-1})$. The solid line is for the *aa* index, the dot-dash line is for the radial component of the IMF near Earth, $|B_{rE}| = F_s/(2\pi R_1^2)$. (bottom). Normalised power spectra density at a period of 0.5 years, $P^*(0.5^{-1})$. In both cases the spectra are determined for 16-year periods about the time shown (as shown in Figure 15). The vertical dashed lines are the times of sunspot minima.

the normalized powers $P^*(1.3^{-1})$ and $P^*(0.5^{-1})$ are the power close to the solid lines in Figure 15, divided by the mean background power in the rest of the band between the dotted lines in the figure.

The top panel of Figure 16 presents an analysis of the variations of this normalized power spectra density at 1.3 (±0.1) years, $P^*(1.3^{-1})$ deduced for 16-year intervals about the time shown on the ordinate axis. The solid line is for the *aa* index, and the dash-dotted line is for the radial component of the IMF measured by spacecraft near Earth, $|B_{rE}| = F_s/(2\pi R_1^2)$. The lower panel shows the power in the semiannual variation, $P^*(0.5^{-1})$. The vertical dashed lines mark the sunspot minima.

The normalized power in the semiannual variation rises and falls with the coronal source flux, as shown in Figure 1, as we would expect because it arises from the rôle of the consequent heliospheric field. However, the bottom panel of Figure 16 shows that there is also an increasingly strong Hale (roughly 22-year period) cycle with clear additional peaks in the declining phases of even-numbered sunspot cycles. As pointed out by *Hapgood* [1993] and by *Cliver et al.* [1996], this is also when the Earth repeatedly intersects more and faster flow streams emerging from the low-latitude extension of coronal holes: These also show a clear 22-year variation.

Both the *aa* index and the interplanetary magnetic field show an increase in the variation at 1.3 years since about 1980. However, the appearance of this variation at earlier times in the *aa* index data is puzzling. The largest powers at this period are seen around the peaks of the even-numbered sunspot cycles, with minima around the peaks of the odd-numbered cycles. Thus this appears to be mainly controlled by the Hale cycle of 22 years, rather than the 11-year sunspot cycle. Furthermore, this behavior is manifest only after about 1940, before which power at this period was generally low: peaks for the maxima of cycles 14 and 16 are entirely absent; however, there is a peak at the maximum of cycle 12.

There is no ready explanation of these variations, and their implications have not yet been explored. However, they do serve to sound a cautionary note about extrapolating solar conditions back in time. It is frequently assumed, explicitly or implicitly, that certain features that are revealed by modern instrumentation have always been present. For example, *Lockwood et al.* [1999a], in their derivation of the coronal source flux variation, assumed that the result reported by *Balogh et al.* [1995] and *Lockwood et al.* [1999a] (namely, that the latitudinal gradients in the radial heliospheric field are small) has been true throughout the 150 years that they studied. We have no evidence to suggest that this was not the case – indeed the good correlations with cosmic rays and the ^{10}Be isotope abundance reported in sections 3 and 4 argues that the assumption is valid. However, Figure 16b cautions that we cannot always assume that we can extrapolate present solar behaviour backward in time.

6. Conclusions and Implications

6.1 Solar Cycle Length.

Recent work has revealed that there have been important long-term (here meaning ~100 year) changes in the magnetic field in the solar atmosphere. Some of this field permeates the coronal source surface and enters the heliosphere. An understanding of these changes is becoming available to us, in terms of the emergence of flux in active regions, its transfer to the network, and the balance between loss of open network flux and its accumulation in coronal holes. In particular, this gives us an understanding of why the length of the solar cycle has important implications. Shorter solar cycles facilitate a rise in the coronal source flux; longer cycles allow it to decay. However, the accumulation of the coronal source flux is also strongly dependent on the rate of flux emergence in active re-

gions. In general, the peak and cycle-averaged sunspot numbers are larger when cycles are shorter. Thus shorter cycles can be associated with larger flux emergence rates. This adds to the effect of the smaller time available for open flux decay. These effects mean that the shorter solar cycles correlate with decreasing coronal source flux.

It is certain that these changes in the coronal source flux mean that the "space climate" in the near-Earth environment will have altered, indeed, the changes in the solar coronal field were detected using the *aa* index of geomagnetic activity. A rise in the coronal source flux increases the coupling of energy between the solar wind flow and Earth's magnetosphere, and the resultant effects on a variety of technological systems will increase, both in number and severity, as F_s rises.

6.2 Cosmic Rays

In addition to influencing the transfer of energy from the solar wind to the magnetosphere, the rise in F_s will have caused the cosmic ray flux incident on the Earth to have fallen. In this paper, we have estimated that the cosmic ray fluxes above 3 GeV were 15% higher, on average, around 1900 than they are now. The corresponding figure for >13 GeV particles is about 4%. The potential implications of this are not yet understood. The correlation between the coronal source flux deduced from geomagnetic activity and the spallation products of cosmic rays found in ice cores and meteorites means that we can use the latter to study variations on longer timescales.

Cosmic rays also have effects on modern technology, and so the long-term variations of their fluxes with F_s are also part of space climate. They may also have some effects on the lower atmosphere and influence the climate at Earth's surface.

In particular, cosmic rays are a key part of the global electric circuit that is driven primarily by thunderstorms [*Bering et al.,* 1998]. They generate air ions in the subionospheric gap which allows current to flow between the tops of thunderclouds and the ionosphere and also between the ionosphere and the ground in association with the fair-weather electric field. It is not known what sort of modulation to this circuit could be brought about by the changes in cosmic ray fluxes shown in Figures 10 and 13.

Svensmark and Friis-Christensen [1997] and *Svensmark* [1998] have reported a solar cycle variation in the global fraction of terrestrial cloud cover seen from space. These authors proposed that this is because cloud cover is directly influenced by the galactic cosmic rays incident on Earth. The cloud data used by *Svensmark and Friis-Christensen* [1997] and *Svensmark* [1998] were compiled from a variety of sources. Concerns about this compilation of diverse and uncalibrated cloud data were pointed out by a number of authors (see review by *Soon et al.* [2000]) and the correlation was made weaker when data for after 1992 became available [*Kristjánsson and Kristiansen*, 2000]. There was also no clear dependence on cloud type, as would have been expected. Another criticism has been that the polar regions were excluded from the study because cloud cover there could not always be distinguished from the ice caps. However, subsequent work has shown that there is a strong correlation for one subset of the cloud cover, as seen at infrared (IR) wavelengths [N. Marsh and H. Svensmark, private communication, 1999; *Marsh and Svensmark*, 2000]. These IR observations are possible at all latitudes (for both day and night), and so true full-globe averages could be used. The data are the "D2" set compiled and inter-calibrated by the International Satellite Cloud Climate Project (ISCCP) [*Rossow et al.*, 1996]. Compiling an homogeneous data series on cloud cover is fraught with difficulties, and *Brest et al* [1997] point out that the solar cycle variations at visible wavelengths reported by *Svensmark* [1998] are small and that the uncertainties in instrument calibration may be a factor, although smaller than the 3% solar cycle variation detected. However, the solar cycle variation reported by *Marsh and Svensmark* [2000] uses the ISCCP-D2 data at IR wavelengths, for which these uncertainties are much reduced. The strong correlation with cosmic ray fluxes is for clouds that are inferred to be at low altitudes and so is associated mainly with stratus and stratocumulus cloud types. The correlations are strongest for maritime regions at mid-latitudes and are poor in areas dominated by other phenomena such as the El Niño-Southern Oscillation (ENSO). The correlations are compelling, especially with the higher energy of cosmic rays (but not so high in energy that the amplitude of the solar cycle variation becomes negligible). The correlation coefficient for 12-month smoothed data exceeds 0.85.

The mechanism or mechanisms that could result in a correlation between cosmic ray fluxes and the global cloud cover are highly controversial and are certainly not yet understood. It is possible that the cloud cover change is a response to the known irradiance changes during the solar cycle and so varies in concert with the cosmic rays, but there is no causal link between them. However, the correlation with cosmic rays appears to be as high as the anticorrelation with the solar irradiance and such an effect has not been predicted by coupled ocean-atmosphere climate models. That there has been a solar cycle variation in global average cloud cover appears now to be beyond any reasonable doubt but demonstrating that this was not a chance occurrence will require data from several more cycles. This variation must be understood. If the mechanism is the direct effect of cosmic rays on cloud occurrence, the implications are genuinely profound for our understanding of global climate change over the past 150 years because of the drift in comic ray fluxes observed (Figure 13).

The waveform of the heliospheric field variations deduced from geomagnetic activity [*Lockwood et al.*, 1999a], from ^{10}Be in ice cores [*Beer et al.*, 1988], and from ^{44}Ti in meteorites [*Bonino et al.*, 1995] are similar. Thus the isotopic data strongly support the long-term drift in the heliospheric field and in cosmic ray fluxes deduced from the *aa* geomagnetic index. The quantitative values of F_s derived from the *aa* index, allow us to calibrate the other two, longer, data sequences in terms of the coronal source flux.

6.3 Solar Irradiance Reconstruction

Lockwood and Stamper [1999] obtained a correlation of the coronal source flux F_s with solar irradiance measurements of 0.84. In fact, they correlated the data from the various monitors separately and used the regression fits to intercalibrate the instruments. If the intercalibration of the instruments by *Fröhlich and Lean* [1998] is adopted, this correlation falls to 0.79 and if the longer sequence of empirically-modelled irradiance estimates are employed, this correlation falls below 0.5 [*Wang et al.,* 2000b]. The coronal source flux F_s is a small fraction of the field emerging through the photosphere, and it is the latter, and its spatial distribution, that is related to the irradiance variations [*Fligge et al.*, 1998].

It must also be stressed that changes in F_s do not necessarily imply changes in either the strength or the distribution of photospheric field, although there are photospheric changes that correlate well with those in F_s, as discussed in section 1.2. That having been said, by applying the correlations between F_s and irradiance (obtained for the last two solar cycles) to the full F_s data sequence, Lockwood and Stamper derived a long-term irradiance variation that is strikingly similar to that of *Lean et al.* [1995]. This agreement is even closer for the revised reconstruction presented recently by *Lean* [2000]. Lean et al. used a method which employs sunspot numbers added to a long-term drift with a waveform given by 11-year running means of sunspot numbers, and with an amplitude based on a comparison of the Maunder minimum and noncyclic Sun-like stars. The extrapolation by Lockwood and Stamper is also quite similar to the reconstruction by *Solanki and Fligge* [1998], who add contributions by the active regions (based on sunspot numbers) to a quiet sun variation (based on cycle length).

Figure 2 offers a simple practical reason as to why these solar irradiance reconstructions are similar. The coronal source flux F_s, which was used as a proxy for total irradiance by *Lockwood and Stamper* [1999], is highly correlated with the 11-year running means of sunspot number, which *Lean et al.* [1995] and *Lean* [2000] use to give the long-term drift. The coronal source flux is also closely related to the cycle length, which was used as a proxy for total irradiance by *Hoyt and Schatten* [1993] and to give the quiet sun variation by *Solanki and Fligge* [1998]. Thus these methods are not as independent as they initially seem, in terms of the waveform of the irradiance variation they predict. However, the amplitude of the long-term drifts derived are also similar, and this is not derived the same way for the three cases.

The work of *Solanki et al.* [2000] may offer the origins of a theoretical explanation of the agreement between the reconstructions of Lean et al. and Solanki and Fligge with the extrapolation of Lockwood and Stamper. The basic equation in the model of Solanki et al. is the continuity of open network flux:

$$dF_s/dt = \gamma E - F_s/\tau_N \quad (1)$$

where E is the rate of flux emergence through the photospheric active regions, τ_N is the time constant for destruction of open network flux and $\gamma = (1 + \tau_t/\tau_a)^{-1}$ and thus depends on the time constants for flux annihilation in active regions τ_a and for transfer of flux from active regions to the network τ_t. *Solanki et al.* [2000] estimate γ to be 0.015 and derive a best fit with a time constant τ_N of 4 years. Because we know F_s and dF_s/dt from the work of *Lockwood et al.* [1999a], we can compute the flux emergence rate E from equation (1) using these values for γ and τ_N. The result is shown in Figure 17a, where E has been normalized by dividing by its average value for cycle 22 ($\langle E \rangle_{22}$). The dashed line shows the 11-year running mean $\langle E \rangle / \langle E \rangle_{22}$. In Figure 17b, this smoothed variation is compared to the correspondingly normalized and smoothed variation of the coronal source flux F_s. These two variations are very similar. The reason for this similarity is that for the time constant $\tau_N = 4$ years derived by Solanki et al., $|dF_s/dt|$ is much less than (F_s/τ_N) on 11-year timescales. Thus, by (1), the long-term variation in E is very similar to that in F_s. Given that E is the rate of flux emergence in active regions, which are associated with both faculae and sunspots, this may be a valid proxy for solar irradiance due to active regions, although we do not yet know to what extent there are long-term "quiet-sun" irradiance variations that arise from the network outside active regions and which may not be related to E. Equation (1) shows that the emergence rate E is proportional to the quantity $\{(F_s/\tau_N) + dF_s/dt\}$, and this would be a better proxy for the solar irradiance if both long-term (~100 year) and solar cycle variations were mainly caused by sunspots and faculae in active regions.

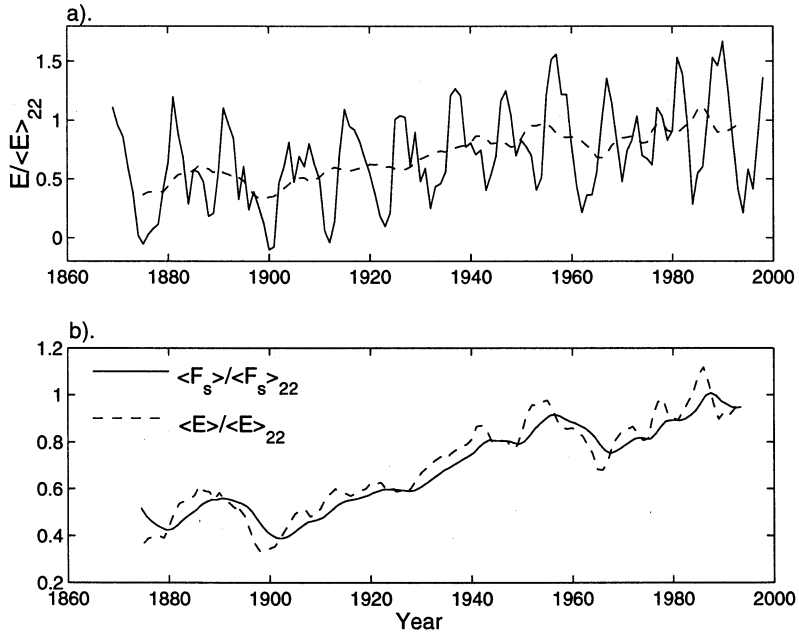

Figure 17. (a) Emergence rate of magnetic flux in active regions, E (solid line) and 11-year running means $\langle E \rangle$ (dashed line), computed from the F_s variation, using equation (1) with a time constant $\tau_N = 4$ years. (b) The 11-year running means $\langle E \rangle$ (dashed line) and $\langle F_s \rangle$ (solid line). All values are normalized with respect to their mean value for cycle 22.

Note added in proof. Since submitting this paper, the author has become aware of a publication by Mursula and Zieger [2000] which reports a 1.3-year oscillation in the *Kp* geomagnetic index since 1932 and also shows that this period appears at the peak of even-numbered sunspot cycles.

Acknowledgments: The author is grateful for discussions with, and pre-prints from, a number of scientists concerning their recent work; in particular, S. Solanki, N. Marsh, H. Svensmark, Y.-M. Wang, G. Cini Castognoli and J. Lean. He is also grateful to J. Beer, who supplied the ^{10}Be isotope ice core data, and to the World Data Center system for collecting, archiving, and distributing the cosmic ray data. This work was supported by the U.K. Particle Physics and Astronomy Research Council.

Janet G. Luhmann thanks both of the referees for their assistance in evaluating this paper.

References

Ahluwalia, H.S., Galactic cosmic ray intensity variations at a high-latitude sea-level site 1937-1994, *J. Geophys. Res., 102*, 24,229-24,236, 1997

Ahluwalia, H.S., and M.D. Wilson, Present status of the recovery phase of cosmic ray 11-year modulation, *J. Geophys. Res., 101*, 4879-4883, 1996.

Balogh, A., E.J. Smith, B.T. Tsurutani, D.J. Southwood, R.J. Forsyth, and T.S. Horbury, The heliospheric field over the south polar region of the Sun, *Science, 268*, 1007-1010, 1995.

Beer, J., S. Tobias, and N. Weiss, An active Sun throughout the Maunder minimum, *Sol. Phys., 181*, 237-249, 1998.

Bering, E.A., A.A. Few, and J.R. Benbrook, The global electric circuit, *Phys. Today, 51*, 24-30, 1998.

Bonino, G., G. Cini Castognoli, N. Bhabdari and C, Taricco, Behavior of the heliosphere over prolonged solar quiet periods by ^{44}Ti measurements in meteorites, *Science, 270,* 1648-1650.

Bradley, R.S., and P.D. Jones, "Little Ice Age" summer temperature variations: Their nature and relevance to recent global warming trends, *Holocene 3,* 367-376, 1993.

Brest, C.L., W.B. Rossow, and M.D. Roiter, Update of radiance callibrations for ISCCP, *J. Atmos. Oceanic Technolo.*, 14, 1091-1109, 1997.

Cane, H.V., G. Wibberenz, I.G. Richardson, and T.T. von Rosenvinge, Cosmic ray modulation and the solar magnetic field, *Geophys. Res. Lett., 26*, 565-568, 1999.

Chapman, G.A., A.M. Cookson, and J.J. Dobias, Solar variability and the relation of facular to sunspot areas during cycle 22, *Astrophys. J.*, 842, 541-545, 1997.

Clilverd, M.A., T.D.G. Clark, E. Clarke, and H. Rishbeth, Increased magnetic storm activity from 1868 to 1995, *J. Atmos. Sol.-Terr. Phys., 60*, 1047-1056, 1998.

Cliver, E.W., V. Boriakoff, and K.H. Bounar, The 22-year cycle of geomagnetic activity, *J. Geophys. Res., 101*, 27,091-27,109, 1996.

Cliver, E.W., V. Boriakoff, and K.H. Bounar, Geomagnetic activity and the solar wind during the Maunder minimum, *Geophys. Res. Lett., 25,* 897-900, 1998a.

Cliver, E.W., V. Boriakoff, and J. Feynman, Solar variability and climate change: Geomagnetic activity and global surface temperature, *Geophys. Res. Lett., 25*, 1035-1038, 1998b.

Cummings, A.C., E.C. Stone, and W.R. Webber, Distance to the solar wind termination shock and the source of anomalous cosmic rays during 1986-1988, *J. Geophys. Res., 99*, 11,547-11,552, 1994.

Feynman, J., and N.U. Crooker, The solar wind at the turn of the century, *Nature, 275*, 626-627, 1978.

Fligge, M., S.K. Solanki, Y.C. Unruh, C. Fröhlich and C. Wehrli, A model of solar total and spectral irradiance variations, *Astron. Astrophys., 335*, 709-718, 1998.

Fligge, M., S.K. Solanki, and J. Beer, Determination of solar cycle length using the continuous wavelet transform, *Astron. Astrophys, 346*, 313, 1999.

Foukal, P., and. J. Lean, An empirical model of total solar irradiance variation between 1874 and 1988, *Science, 247*, 556-558, 1990

Friis-Christensen, E., and K. Lassen, The length of the solar cycle, an indicator of solar activity closely associated with climate, *Science, 192*, 1189-1202, 1991.

Fröhlich, C., and J. Lean, The Sun's total irradiance: Cycles, trends and related climate change uncertainties since 1976, *Geophys. Res. Lett., 25*, 4377-4380, 1998.

Gazis, P.R., Solar cycle variation of the heliosphere, *Rev. Geophys., 34,* 379-402, 1996.

Gazis, P.R., J.D. Richardson, and K.I. Paularena, Long-term periodicity in the solar wind velocity during the last three solar cycles, *Geophys. Res. Lett., 22*, 1165-1168, 1995.

Gleissberg, W., A table of secular variations of the solar cycle, *J. Geophys. Res., 49,* 243-244, 1944.

Hapgood, M.A., A double solar-cycle variation in the 27-day recurrence of geomagnetic activity, *Ann. Geophys., 11*, 248-256, 1993.

Harvey, K.L., and C. Zwaan, Properties and emergence of bipolar active regions, *Sol.* Phys.,148, 85-118, 1993.

Howe, R., J. Christensen-Dalsgaard, F. Hill, R.W. Komm, R.M. Lassen, J. Shou, M.J. Thompson, and J. Toomre, Dynamic variations at the base of the solar convection zone, *Science, 287*, 2456-2460, 2000.

Hoyt, D., and K. Schatten, A discussion of plausible solar irradiance variations 1700-1992, 18,895-18,906, 1993.

Kristjánsson, J.E., and J. Kristiansen, Is there a cosmic ray signal in recent variations in global cloudiness and cloud radiative forcing?, *J. Geophys. Res., 105*, 11,851-11,863, 2000.

Lean, J., Evolution of the Sun's spectral irradiance since the Maunder minimum, *Geophys. Res. Lett., in press*, 2000.

Lean, J., J. Beer, and R. Bradley, Reconstruction of solar irradiance since 1610: Implications for climate change, *Geophys. Res. Lett., 22*, 3195-3198, 1995.

Lockwood, M., and R. Stamper, Long-term drift of the coronal source magnetic flux and the total solar irradiance, *Geophys Res. Lett., 26*, 2461-2464, 1999.

Lockwood, M., R. Stamper, and M.N. Wild, A doubling of the sun's coronal magnetic field during the last 100 years, *Nature*, 399, 437-439, 1999a.

Lockwood, M., R. Stamper, M.N. Wild, A. Balogh, and G. Jones, Our changing Sun, *Astron. Geophys, 40*, 4.10-4.16, 1999b.

Marsh, N., and H. Svensmark, Cosmic rays, clouds and climate, *Space Sci. Rev.*, in press, 2000.

Mayaud, P.N., The *aa* indices: A 100-year series characterising the magnetic activity, *J. Geophys. Res., 77*, 6870-6874, 1972.

Moraal, H., Cosmic ray modulation studies in the outer heliosphere, *Nucl. Phys. B, 33A,B,, proc. suppl.*, 161-178, 1993.

Mursula, K., and T. Ulich, A new method to determine the solar cycle length, *Geophys. Res. Lett., 25*, 1837-1840, 1998.

Mursula, K and B. Zieger, The 1.3-year variation in solar wind speed and geomagnetic activity, *Adv. Space Res.* , in press, 2000.

O'Brien, K., Secular variations in the production of cosmogenic isotopes in the Earth's atmosphere, J. Geophys. Res., 84, 423-431, 1979.

Parker, D.E., P.D. Jones, A. Bevan, and C.K. Folland, Interdecadal changes of surface temperatures since the late 19th century, *J. Geophys. Res., 99*, 14,373-14,399, 1994.

Paularena, K.I., A. Sazbo, and J.D. Richardson, Coincident 1.3-year periodicities in the Ap geomagnetic index and the solar wind, *Geophys. Res. Lett., 22*, 3001-3004, 1995.

Potgieter, M.S., The long-term modulation of galactic cosmic raysnin the heliosphere, *Adv. in Space Res, 16(9),* 191-203, 1995

Pulkkinen, T.I., H. Nevanlinna, P.J. Pulkkinen, and M. Lockwood, The Earth-Sun connection in time scales from years to decades to centuries, *Space Sci. Rev.*, in press, 2000.

Richardson, J.D., K.I. Paularena, J.W. Belcher, and A.J. Lazarus, Solar wind oscillations with a 1.3 year period, *Geophys. Res. Lett., 21*, 1559-1560, 1994.

Rossow, W.B., A.W. Walker, D.E. Beuschel, and M.D. Roiter, International Satellite Cloud Climatology Project (ISCCP): Documentation of new datasets, WMO/TD 737, World Meteorol. Organ., Geneva, 1996.

Russell, C.T., and R.L. McPherron, The magnetotail and substorms, *Space Sci. Rev., 15,* 205-225, 1973.

Sabbah, I., The role of interplanetary magnetic field and solar wind in modulating both galactic cosmic rays and geomagnetic activity, *Geophys. Res. Lett., 27*, 1823-1826, 2000.

Schatten, K.H., J.M. Wilcox, and N.F. Ness, A model of interplanetary and coronal magnetic fields, *Sol. Phys., 6*, 442-455, 1969

Silverman, S.M., and R. Shapiro, Power spectral analysis of auroral occurrence frequency, *J. Geophys. Res., 88*, 6310-6316, 1983.

Solanki , S.K. and M. Fligge, Solar irradiance since 1874 revisited, *Geophys. Res. Lett., 25*, 341-344, 1998.

Solanki, S.K. and M. Fligge, A reconstruction of total solar irradiance since 1700, *Geophys. Res. Lett., 26*, 2465-2468, 1999.

Solanki , S.K., M. Schüssler, and M. Fligge, Secular evolution of the Sun's magnetic field since the Maunder minimum, *Nature*, in press, 2000.

Soon, W., S. Baluinas, E.S.Posmentier, and P. Okcke, Variations of solar coronal hole qarea and terrestrial lower tropospheric air temperature from 1979 to mid-1998: Astronomical forcings of changes in Earth's climate?, New Astron., *4*, 563-579, 2000.

Stamper, R., M. Lockwood, M.N. Wild, and T.D.G. Clark, Solar causes of the long term increase in geomagnetic activity, *J. Geophys Res., 104*, 28,325-28,342, 1999.

Stuiver, M., and P.D. Quay, Changes in atmospheric carbon-14 attributed to a variable Sun, *Science, 207*, 11, 1980.

Svensmark, H., Influence of cosmic rays on Earth's climate, *Phys. Rev. Lett., 81*, 5027-5030, 1998.

Svensmark, H., and E. Friis-Christensen, Variation of cosmic ray flux and global cloud coverage: A missing link in solar-climate relationships, *J. Atmos. Sol. Terr. Phys., 59*, 1225-1232, 1997.

Szabo, A., R.P. Lepping, and J.H. King, Magnetic field observations of the 1.3 solar wind oscillations, *Geophys. Res. Lett., 22*, 1845-1848, 1995.

Tett, S., P.A. Stott, M.R. Allen, W.J. Ingram and J.F.B. Mitchell, Causes of twentieth century temperature change, *Nature,388*, 569-572, 1999.

Usoskin, I.G., H. Kananen, K. Mursula, P. Tanskanen, and G.A. Kovaltsov, Correlative study of solar activity and cosmic ray intensity, *J. Geophys.Res., 103*, 9567-9574, 1998.

Vasyliunas, V.M., J.R. Kan, G.L. Siscoe, and S.-I. Akasofu, Scaling relations governing magnetospheric energy transfer, *Planet Space Sci., 30*, 359-365, 1982.

Wang, Y.-M. and N.R. Sheeley Jr., Solar implications of Ulysses interplanetary field measurements, *Astrophys. J., 447*, L143-L146, 1995.

Wang, Y.-M., S.H. Hawley, and N.R. Sheeley Jr.,. The magnetic nature of coronal holes, *Science, 271*, 464-469, 1996.

Wang, Y.-M., N.R. Sheeley Jr., and J. Lean, Understanding the evolution of Sun's magnetic flux, *Geophys. Res. Lett., 27*, 621-624, 2000a.

Wang , Y.-M., J. Lean, and N.R. Sheeley Jr., The long-term evolution of the Sun's open magnetic flux, *Geophys. Res. Lett., 27*, 505-508, 2000b.

Wigley, T.M.L., and S.C.B. Raper, Climatic change due to solar irradiance changes, *Geophys. Res. Lett., 17*, 2169-2172, 1990.

Willson, R.C., Total solar irradiance trend during cycles 21 and 22, *Science, 277*, 1963-1965, 1997.

M. Lockwood, World Data Centre C-1 for Solar-Terrestrial Physics, Space Science Department, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, England, U.K. (m.lockwood@rl.ac.uk)

(Received May 16, 2000; revised June 28, 2000;
accepted August 2, 2000.)